bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sanctions, partner recognition, and variation in mutualistic symbiosis

Jeremy B. Yoder and Peter Tiffin

Dept. of Forest & Conservation Sciences, University of British Columbia, Vancouver, BC, Canada V6T 1Z4; jbyoder@mail.ubc.ca

Dept. of Plant Biology, University of Minnesota, Saint Paul, MN, USA 55108

20 April 2016

Abstract

Most mechanisms proposed to protect mutualistic species interactions against invasion by non-cooperative individuals imply that mutualist populations should consist only of fully cooperative individuals. Empirical studies, however, find that genetic variation in mutualist quality is widespread. A possible explanation for this paradox is that mutualistic interactions are mediated not only by assessment of partner quality, but also by genetically independent partner signals. Here, we build and evaluate a model of host-symbiont coevolution in which alleles at one locus determine host recognition of compatible symbionts, and alleles at a second locus determine whether hosts are able to sanction non-cooperative symbionts. Symbionts' expression of signals recognized by the hosts and cooperation in mutualism are similarly determined by separate loci. In contrast to simpler models that either fail to maintain variation in symbiosis outcomes or long-term persistence of cooperative symbionts, this model can maintain variation in both species. Individual-based simulations conducted over a range of reasonable parameter values and starting conditions show that the dual-system model is more likely to maintain variation in mutualism outcomes than simpler models, more likely to maintain variation in host loci, and able to promote greater among-site variation in symbiont quality. The

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dual genetic systems of sanctions and host-symbiont recognition also converge toward conditions similar to one recently-developed model of symbiosis, in which hosts offering the right incentives to potential symbionts can initiate symbiosis without prior screening for partner quality. Our results suggest that a full understanding of mutualistic symbiosis requires integration of communication between partner species as well as the exchange of benefits.

Keywords: *symbiosis, mutualism, coevolution, population genetics*

Introduction

Mutually beneficial interactions between species pose two related conundrums for evolutionary biology. First, how are mutualisms maintained in the face of the apparent advantages to individuals who accept resources or services but provide none in return? And second, given a mechanism that prevents the evolution of non-cooperative participants, why do members of interacting species vary in mutualistic quality?

Evolutionary theory offers multiple solutions to the first conundrum, in the form of coevolutionary dynamics that ensure non-cooperators are at a net fitness disadvantage over the long term, even if they have a short-term advantage over cooperators. Partner choice allows individuals to discontinue interaction with non-cooperators (Trivers 1971; Axelrod and Hamilton 1981; Foster et al. 2006), or sanction non-cooperators by providing reduced rewards (Bull and Rice 1991; West et al. 2002a; West et al. 2002b; Sachs et al. 2004; Akçay and Simms 2011). Similar to this is partner fidelity feedback, in which cooperation prompts physiological responses such that cooperative partners receive greater rewards without any active "decision" by the reward-providing species (Doebeli and Knowlton 1998; Weyl et al. 2010; Archetti and Scheuring 2011).

Many real-world mutualistic interactions do appear to work like this. Soybean plants are capable of cutting off support to root nodules containing rhizobial bacteria that do not produce bioavailable nitrogen (Kiers et al. 2003) and can scale these sanctions quantitatively to reduce support for less-productive nodules (Kiers et al. 2006). In the obligate pollination mutualisms of yuccas and figs, host plants' floral abortion responses prevent over-exploitation by

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

seed-feeding pollinators (Pellmyr and Huth 1994; Jandér and Herre 2010). Reduced growth of ant domatia on herbivore-damaged branches of ant-protected shrubs has also been described as a possible sanctioning response to poor protection by ants (Edwards et al. 2006).

The second conundrum arises from these solutions to the first. In the absence of a cost to the species that exercises them, partner choice, sanctions, and partner fidelity feedback can all lead to fixation of cooperative genotypes (Axelrod and Hamilton 1981; Doebeli and Knowlton 1998; West et al. 2002a; West et al. 2002b). Similarly, models that assume mutualists maximize their fitness when their traits are matched predict reduced diversity in each of the interacting species (Kiester et al. 1984; Kopp and Gavrilets 2006; Yoder and Nuismer 2010). Yet genetic variation in partner quality is widely observed in natural populations of mutualists (Heath and Stinchcombe 2013), including rhizobial bacteria (Simms and Taylor 2002; Heath et al. 2010) and mycorrhizal fungi (Hoeksema 2010), ant bodyguards (Ness et al. 2006), and seed-feeding obligate pollinators (Pellmyr and Huth 1994; Herre and West 1997; Holland et al. 1999). Indeed, such variation creates the selection necessary to maintain costly countermeasures against non-cooperative partners (Foster et al. 2006; McNamara and Leimar 2010), and interactions with non-cooperators can maintain higher density of cooperators (Jones et al. 2009).

In contrast to models of mutualism, models of antagonistic interactions provide coevolutionary dynamics that maintain genetic variation in interacting species, either through by creating inverse frequency dependent selection on one or both partners (e.g., Dieckmann et al. 1995; Agrawal and Lively 2002), or other fitness outcomes that depend on phenotypic matching (Sasaki 2000; Nuismer and Otto 2005; Kopp and Gavrilets 2006; Yoder and Nuismer 2010). Such models are generally drawn from biological systems in which host defensive responses are activated by recognition of molecules expressed by parasites or pathogens (Dybdahl et al. 2014; Nuismer and Dybdahl 2016).

Similar recognition may play a role in mutualism, though it is not often incorporated into models of such interactions. Legumes exchange and rhizobia exchange molecular signals as they establish symbiosis (Triplett and Sadowsky 1992), and gene families associated with pathogen recognition have also been implicated in legume-rhizobium compatibility (Yang et al. 2010; Young et al. 2011). Interestingly, host-rhizobium compatibility is at least partly independent of variation in mutualism outcomes (Triplett and Sadowsky

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1992; Heath 2009). Multiple obligate pollination mutualisms are mediated by complex, species-specific floral scents that are presumably independent of rewards offered to pollinators (Svensson et al. 2005; Okamoto et al. 2007; Soler et al. 2011). Similarly, host plant volatiles guide the colonizing queens of plant-protecting ant species and direct the activity of ants' patrols (Edwards et al. 2006; Edwards et al. 2007; Schatz et al. 2009), though the chemical basis of these volatiles can be simple and non-specific (Schatz et al. 2009). Immune recognition responses also mediate the assembly of animals' microbiomes (Cullender et al. 2013; Mutlu et al. 2014; Fishman 2015), which may have substantial impacts on host health (Cho and Blaser 2012; Pflughoeft and Versalovic 2012).

These interspecific signals demonstrate that successful mutualistic interactions rely on more than the quality of goods or services exchanged; communication between potential partners also is necessary to attract and establish mutualisms. Because signaling and benefits are likely independent traits in most mutualisms, a model incorporating independent, coevolving genetic systems for signal exchange and mutualist performance might better capture the complexity of empirical systems. Here, we develop and evaluate models of host-symbiont mutualism mediated by (1) host sanctions against non-cooperative symbionts, (2) host recognition of symbiont signals that are independent of symbiont quality, and (3) host sanctioning of non-cooperators paired with recognition of symbionts.

After developing analytic models of allele frequency dynamics within a population of hosts and symbionts, we use individual-based coevolutionary simulations to examine a wider range of parameters, and to incorporate the effects of genetic drift across a metapopulation of sites connected by migration. We show that the model incorporating both recogition and sanctions is a better fit for empirical observations of variation in mutualism outcome than models with only sanctions or partner recognition alone. Evaluation of this model also reveals conditions under which mutualisms should evolve as predicted by simpler models of host-symbiont interaction.

Analytic Models

In all three models, we consider the interaction of haploid hosts and symbionts and track allele frequencies in both species within a single population. We

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

assume that hosts and symbionts encounter each other at random, and that each receives a benefit B and pays a cost C of interaction. For both host and symbiont we assume that fitness is equal to $1 + P$, where P is the payout (i.e., net benefit) from the interaction with the other species. Payout is determined by host and symbiont genotype, by the possible benefit (B_i) and the cost (C_i) of interaction to each species i, and by the nature of the host-symbiont interaction.

Host sanctions

First, consider a model of host sanctions against non-cooperative symbionts, with symbiont cooperation and the host's ability to sanction each determined by a single biallelic locus. Hosts and symbionts meet at random, and engage in symbiosis. Symbionts with the M allele at a *cooperation* locus cooperate with the host; symbionts with the m allele do not. Interacting with a cooperative symbiont, all hosts pay a cost of hosting symbionts, C_H, and receive a benefit of symbiosis, B_H; cooperative symbionts pay a cost of symbiosis, C_S, and receive a benefit, B_S, while non-cooperative symbionts pay no cost and receive only the benefit.

Interacting with a non-cooperative symbiont, hosts with the H allele at a *sanctions* locus are able to stop the interaction with probability ω; but hosts with the h allele are not able to do so. The term ω determines the probability that sanctioning hosts are able to avoid paying the costs of hosting non-cooperating symbionts. If $\omega = 1$, sanctioning hosts never suffer the cost of hosting non-cooperators; if $\omega = 0$, sanctions have no effect. This approximates a "tit-for-tat" strategy against defection in an iterated Prisoner's Dilemma game (Axelrod and Hamilton 1981; Ohtsuki 2010).

As noted above, host and symbiont fitness are equal to $1 + P$, where P is the payout from the interaction with the other species, determined by host and symbiont genotype, by the possible benefit (B_i) and the cost (C_i) of interaction to each species i, and, in models with host sanctions, the effectiveness of sanctions. From this, we can derive the payouts associated with interactions between all possible combinations of host and symbiont genotypes (Table 1).

From these payouts, we can calculate the fitness of each host and symbiont genotype given the frequency of the symbiont cooperation allele, p_M, and

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: Host and symbiont payouts under the model of host sanctions.

	Host	
Symbiont	H	h
	Host payout	
M	$B_H - C_H$	$B_H - C_H$
m	$(1-\omega)C_H$	$-C_H$
	Symbiont payout	
M	$B_S - C_S$	$B_S - C_S$
m	$(1-\omega)B_S$	B_S

the host sanctioning allele, p_H, and derive the per-generation change in the frequency of the host's H allele:

$$\Delta p_H = p_H(1-p_H)\frac{\omega C_H(1-p_M)}{1 - B_H p_M - C_H[1-\omega p_H(1-p_M)]} \tag{1}$$

And the symbiont's M allele:

$$\Delta p_M = p_M(1-p_M)\frac{\omega B_S p_H - C_S}{1 + B_S[1-\omega p_H(1-p_M)] - C_S p_M} \tag{2}$$

Partner recognition

As an alternative, consider a model of partner recognition in which hosts only interact with symbionts expressing a signal that is determined by a locus unlinked to the locus that determines whether symbionts cooperate.

As in the previous model, symbionts cooperate if they have the M allele at the cooperation locus, and do not if they have the m allele; but they also carry either a S allele or a s allele at a second *signaling* locus. If hosts carry the R allele at a *recogition* locus, they initiate symbiosis only with symbionts carrying the S signaling allele, and if they carry the r allele, they initiate symbiosis only with symbionts carrying s. Hosts have no ability to sanction non-cooperating symbionts; host payouts are determined solely by whether

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2: Host and symbiont payouts under the model of partner recognition.

Symbiont	Host R	r
	Host payout	
MS	$B_H - C_H$	0
Ms	0	$B_H - C_H$
mS	$-C_H$	0
ms	0	$-C_H$
	Symbiont payout	
MS	$B_S - C_S$	0
Ms	0	$B_S - C_S$
mS	B_S	0
ms	0	B_S

symbionts with compatible recognition alleles are also cooperative. This results in the payouts for host-symbiont genotype pairings shown in Table 2.

An exact analytic examination of equilibria in this model is impractical. However, if we assume that the costs and benefits of the interaction are small (Nuismer et al. 2010; Yoder and Nuismer 2010), that the effects of the symbiont cooperation (M) and signaling (S) loci are therefore not strongly epistatic, and that there is free recombination between symbiont loci, then LD between these loci should approach quasi-linkage equilibrium (QLE) conditions (Barton and Turelli 1991; Kirkpatrick et al. 2002). If these conditions hold, we can approximate change in the frequency of the host R allele as

$$\Delta p_R \approx p_R(1 - p_R)(B_H p_M - C_H)(2p_S - 1) \tag{3}$$

We can similarly approximate change in the frequency of the symbiont's M allele

$$\Delta p_M \approx p_M(1 - p_M)\left[p_S - p_R(2p_S - 1) - 1\right] C_S \tag{4}$$

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

And in the frequency of the S allele

$$\Delta p_S \approx p_S(1-p_S)(2p_R-1)(B_S - C_S p_M) \tag{5}$$

Finally, we can approximate change in linkage disequilibrium (LD) between the symbiont's mutualism and recognition loci, δ_S:

$$\Delta\delta_S \approx -\frac{1}{2}\left[p_M(1-p_M)p_S(1-p_S)(2p_R-1)C_S + \delta_S\right] \tag{6}$$

Note the approximations for change in allele frequencies do not refer to the LD terms δ_S — as a conseqeunce of the assumption that LD is weak, changes in allele frequency at each symbiont locus are independent. Also note that $\Delta\delta_S$ has the opposite sign of δ_S, which means that LD between the symbiont cooperation and signaling loci will evolve toward zero, unless $-p_M(1-p_M)p_S(1-p_S)(2p_R-1)C_S < \delta_S < 0$, or $0 < \delta_S < -p_M(1-p_M)p_S(1-p_S)(2p_R-1)C_S$. For parameter values meeting the assumptions of the approximation (small C_S) this means that LD between symbiont loci will remain negligible.

Recognition with sanctions

Finally, we consider a model in which hosts have loci for both symbiont recognition and sanctions of non-cooperative symbionts, and symbionts have both signaling and cooperation loci as described in the partner recognition model. Hosts initiate symbiosis only with symbionts carrying a signalling allele compatible with the hosts' genotype at the recognition locus, as in the partner recognition model. However, hosts are also able to sanction if they carry the H allele at the sanctioning locus, as in the host sanctions model.

As in the host recognition model, in order to develop a tractable model we approximate assuming that the costs and benefits of interaction are small, and LD is low between freely recombining host and symbiote loci. The payoff values for each possible combination of host and symbiont genotypes (Table 3) then lead to the following approximations of change in the allele frequency at each locus, and change in LD between loci in each species. For the host, these are

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3: Host and symbiont payouts under the model of recognition with sanctions.

Symbiont	Host HR	Hr	hR	hr
	Host payout			
MS	$B_H - C_H$	0	$B_H - C_H$	0
Ms	0	$B_H - C_H$	0	$B_H - C_H$
mS	$-(1-\omega)C_H$	0	$-C_H$	0
ms	0	$-(1-\omega)C_H$	0	$-C_H$
	Symbiont payout			
MS	$B_S - C_S$	0	$B_S - C_S$	0
Ms	0	$B_S - C_S$	0	$B_S - C_S$
mS	$(1-\omega)B_S$	0	B_S	0
ms	0	$(1-\omega)B_S$	0	B_S

$$\begin{aligned}\Delta p_H \approx\ & p_H(1-p_H)(1-p_M) \\ & [1 - p_S - p_R(2p_S - 1)]\,\omega C_H\end{aligned} \tag{7}$$

$$\begin{aligned}\Delta p_R \approx\ & p_R(1-p_R)(2p_S - 1) \\ & [B_H p_M - p_H(1-p_M)\omega C_H - C_H]\end{aligned} \tag{8}$$

$$\begin{aligned}\Delta \delta_H \approx\ & -\frac{1}{2}[p_H(1-p_H)p_R(1-p_R) \\ & (1-p_M)(2p_S-1)\omega C_H + \delta_H]\end{aligned} \tag{9}$$

And, for the symbiont

$$\begin{aligned}\Delta p_M \approx\ & p_M(1-p_M)\left[p_R p_S - C_S(p_H(1-p_R)(1-p_S)\right] - \\ & B_S\left\{p_R(1-p_S) - p_H\left[1 - p_S - p_R(1-\omega-p_S)\right]\right\}\end{aligned} \tag{10}$$

$$\begin{aligned}\Delta p_S \approx\ & p_S(1-p_S)p_M \\ & \left[p_R - p_H(1-p_R)\right](B_S - C_S)\end{aligned} \tag{11}$$

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\Delta\delta_S \approx \frac{1}{2}\{p_M(1-p_M)p_S(1-p_S) \\ [p_R - p_H(1-p_R)]\,(B_S - C_S) - \delta_S\} \tag{12}$$

As in the host-symbiont recognition model, the approximations for change in allele frequencies do not include the terms δ_H and δ_S, meaning that LD between loci does not contribute to the approximated change in allele frequencies for either species. For the hosts, $\Delta\delta_H$ and δ_H have opposite signs, and LD between the sanctions and recognition loci evolves toward zero, unless $-p_H(1-p_H)p_R(1-p_R)(1-p_M)(2p_S-1)\omega C_H < \delta_H < 0$ or $0 < \delta_H < -p_H(1-p_H)p_R(1-p_R)(1-p_M)(2p_S-1)\omega C_H$. In symbionts, LD between the cooperation and signaling loci evolves toward zero unless $-p_M(1-p_M)p_S(1-p_S)[p_R - p_H(1-p_R)](B_S - C_S) < \delta_S < 0$ or $0 < \delta_S < -p_M(1-p_M)p_S(1-p_S)[p_R - p_H(1-p_R)](B_S - C_S)$. In both species, the conditions required for the approximation (small cost, C_i and benefit, B_i terms for each species) make the values of LD in these ranges negligibly small.

Model evaluation

For each of the three analytic models, we solve for equilibria in the models derived above. We are particularly interested in equilibria that maintain intermediate frequencies of mutualistic symbionts (i.e., $0 < p_M < 1$), indicating conditions under which coevolutionary selection permits genetic variation in mutualist quality (Heath and Stinchcombe 2013); but genetic variation in the outcomes of symbiosis is also implied by equilibria with intermediate frequencies at the host recognition and symbiont signaling loci. Full details of these analyses are given in Mathematica notebooks, available at github.com/jbyoder/mutualism-sanctions-recognition.

Host sanctions

The model of host sanctions in a single population has equilibria when $p_M = 0$ and p_H is equal to either 1 or 0; and when $p_M = 1$ at any value of p_H. All of these can be locally stable given the right parameter values. The rate and direction of change in allele frequencies in this model suggest that

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

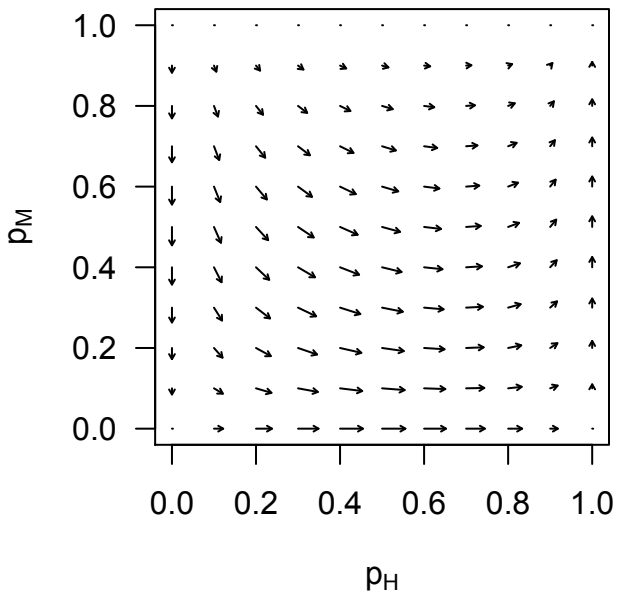

Figure 1: Dynamics of the host sanctions model. Vector-field plot indicating magnitude and dirction of change in the frequency of host sanctions (p_H) and symbiont cooperation (p_M) alleles at given starting frequencies, with $C_H = C_S = 0.25$, $B_H = B_S = 0.5$, and $\omega = 0.75$.

mutualistic symbionts will be maintained near fixation over the long term when $\omega B_S > C_S$ (Figure 1). If $C_S > \omega B_S p_H$, cooperative symbionts have lower fitness than non-cooperators, and the M allele becomes less common. However, sanctioning hosts have higher fitness than non-sanctioning hosts whenever $p_M < 1$, provided sanctions are effective ($\omega > 0$) and there is a cost to hosting symbionts ($C_H > 0$). This fitness advantage increases as cooperative symbionts become less common, which increases p_H until cooperative symbionts have higher fitness (when $p_H > \frac{C_S}{\omega B_S}$), leading to fixation of the M allele.

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

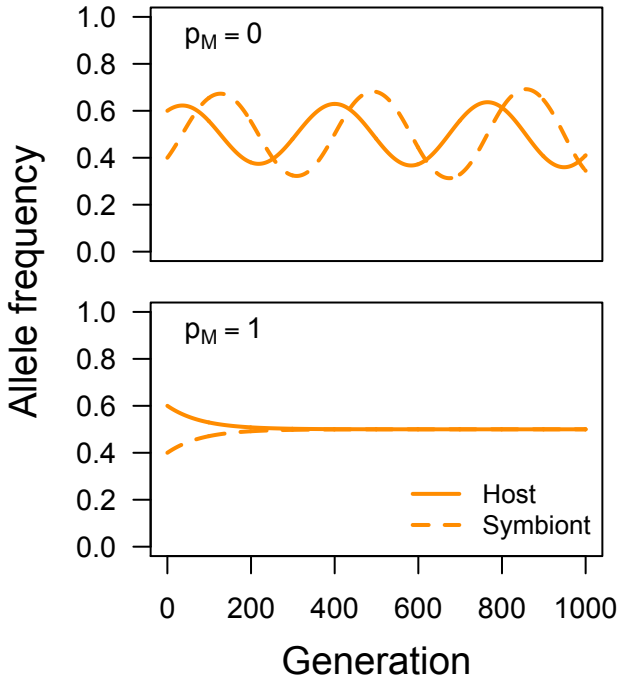

Figure 2: Dynamics of partner recognition. Allele frequencies over time at host (solid lines) and symbiont (dashed lines) recognition loci, with intial frequency of the symbiont cooperation allele $p_M = 0$ (top) or $p_M = 1$ (bottom), and with $C_H = C_S = 0.025$, $B_H = B_S = 0.05$, and $\delta = 0.1$.

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Partner recognition

Perhaps the most important feature of the model of partner recognition is that its approximation of change in the frequency of the symbiont cooperation allele Δp_M, is necessarily less than or equal to zero for all reasonable parameter values (see derivation in the "recognition" Mathematica notebook at github.com/jbyoder/mutualism-sanctions-recognition). That is, there is no condition under which a rare cooperative symbiont genotype would increase in frequency until fixation. This means that the system can be fairly described as a mutualism ($p_M > 0$) only if the cooperation allele is already present at non-zero frequency in the population, and conditions prevent its loss (i.e., $\Delta p_M = 0$, its maximum value). There are multiple such coevolutionary equilbria for the partner recognition system, including some where $p_M > 0$.

Equilibria exist at which variation is maintained either at host and symbiont recognition loci, or at the symbiont cooperation locus, but none are locally stable. Two of these occur when $p_R = p_S = \frac{1}{2}$, and p_M is equal to either 0 or 1 — that is, cooperative symbionts are either lost or fixed. In the first case, when $p_M = 0$, oscillations occur because hosts experience negative frequency-dependent selection for the recognition genotype that is compatible with the least-common symbiont type (Figure 2, top). That is, when mutualistic symbionts are absent, hosts maximize their fitness by avoiding symbiosis altogether. In the second case, equilibrium occurs because all symbionts are cooperative, so there is no difference in fitness between hosts with different recognition alleles. At this equilibrium, there is no variation in the outcome of symbiosis (Figure 2, bottom).

Unstable equilibria also occur when $p_M = \frac{C_S}{B_S}$, and when the host recognition locus and symbiont signaling locus are fixed for incompatible alleles (i.e., $p_R = 1$ and $p_S = 0$, or $p_R = 0$ and $p_S = 1$). At these equilibria, symbionts maintain variation for cooperation, but host and symbiont recognition genotypes ensure that symbiosis is never initiated.

There is no multivariable equilibrium at which LD between the symbiont's M and S loci, δ_S, is greater than zero. That is, coevolutionary selection in this model cannot create a stable association between recognition alleles and mutualism alleles, in spite of the advantage for hosts carrying the recognition allele more strongly associated with the cooperation allele in symbionts. This may reflect the constraints of the quasi-linkage-equilibrium, weak-selection

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

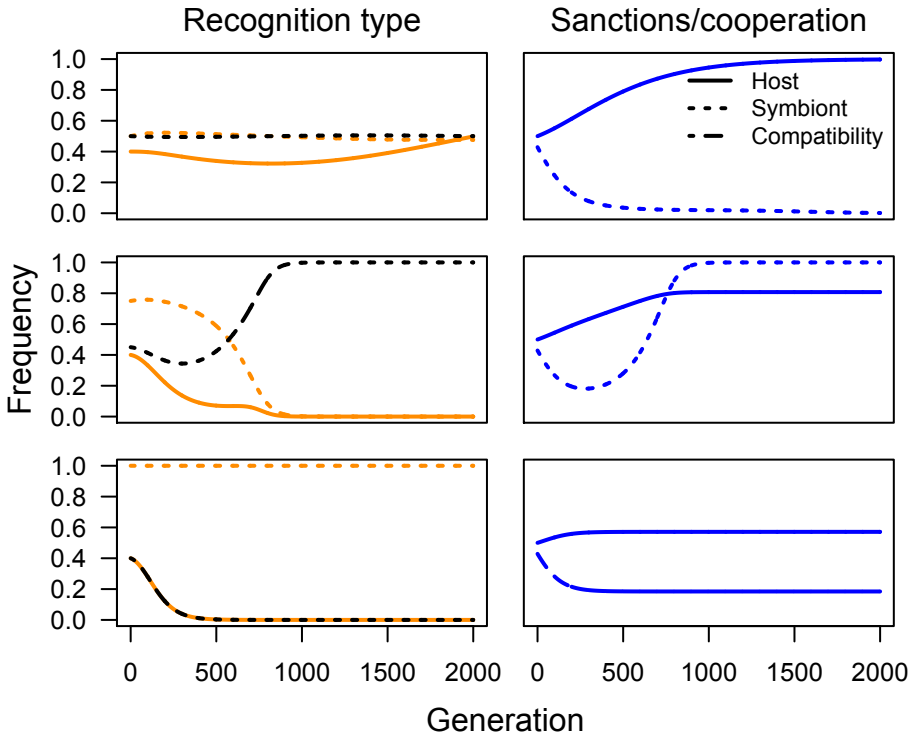

Figure 3: Dynamics of partner recognition and host sanctions. Allele frequencies over time at host (solid lines) and symbiont (dashed lines) recognition loci (left, orange) or sanctions/cooperation loci (right, blue), when the intial frequency of the symbiont recognition allele $p_{S0} = 0.5$ (top), $p_{S0} = 0.75$ (middle), or $p_{S0} = 1$ (bottom). In plots of allele frequency at the recognition loci (left), a dotted black line indicates the probability that a randomly-drawn host and symbiont will have compatible recognition genotypes. In all panels, $C_H = C_S = 0.025$, $B_H = B_S = 0.05$, and $\omega = 0.25$.

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

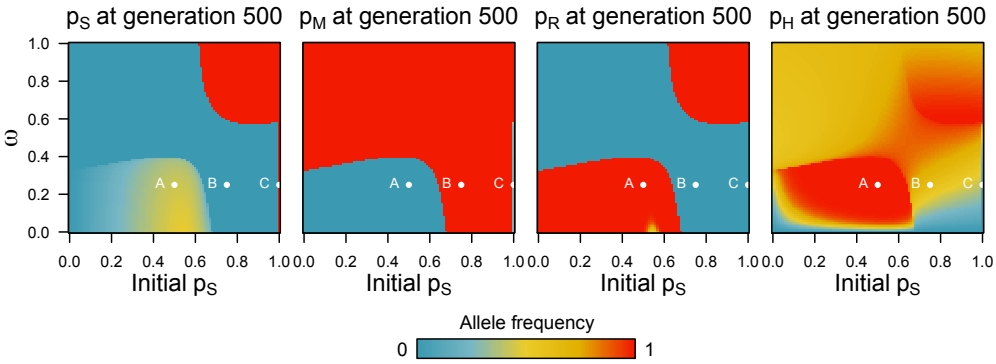

Figure 4: Outcomes of partner recognition and host sanctions. Frequency of alleles at each symbiont and host locus after 500 generations of iterating equations 7-12 with varying sanction effectiveness (ω) and initial frequencies of the symbiont recognition allele (p_S). For all scenarios, the initial frequency of the host recognition allele $p_R = 0.3$, host sanctions $p_H = 0.5$, symbiont cooperation $p_M = 0.5$, $C_H = C_S = 0.25$, $B_H = B_S = 0.5$, and $\delta_H = \delta_S = 0$. White points correspond to coevolutionary recursions illustrated in Figure 3, top (A), middle (B), and bottom (C) panels.

assumptions necessary to calculate equilibria for the system — the individual-based simulations, below, allow relaxation of these assumptions.

Recognition with sanctions

The model of partner recognition with host sanctions has a number of equilibria at which host or symbiont recognition loci remain variable. However, there is no locally stable equilibrium at which hosts and symbionts are not fixed for incompatible alleles and variation is maintained at one or more host or symbiont locus, or at which LD between loci is greater than zero for either hosts or symbionts (i.e., at which the signal-recognition and cooperation-sanction systems evolve non-independently). Iterating the system of generational recursions shows that outcomes of the model can be dependent on starting allele frequencies and the effectiveness of host sanctions, ω.

At sufficiently low values of ω, the system can maintain variation at some loci if other loci become fixed (Figure 3). These dynamics might be described as a

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

race between the recognition/signaling system and the sanctions/cooperation system. If host and symbiont recognition loci start at similar intermediate frequencies (Figure 3, top panel) — so that any given host is compatible with about half of the symbiont population — non-cooperating symbionts have an advantage and symbiont cooperation is lost (dashed blue line). This occurs even as the frequency of sanctioning hosts increases (solid blue line); once sanctioning hosts have fixed, cooperative symbionts are so rare that there is a net cost to initiating symbiosis even given the ability to sanction, and the host type compatible with the smaller fraction of the symbiont population fixes (solid orange line). As the frequency of one host recognition allele approaches fixation, initiating symbiosis becomes so unlikely that symbionts have no opportunity to obtain the benefit of cooperation, and cooperative symbionts are lost from the population.

Shifting the initial frequency of symbiont cooperation alleles, however, creates a situation in which fewer hosts initiate symbiosis. At first, hosts evolve even lower compatibility (Figure 3, middle, solid orange line) and more frequent sanctions (solid blue line), and symbionts evolve lower cooperation (blue dashed line). Nevertheless, host sanctions become sufficiently common to give cooperative symbionts an advantage over non-cooperators before host-symbiont compatibility is entirely lost. Cooperative symbionts then increase in frequency until they are fixed (blue dashed line). When cooperation becomes fixed in the the symbiont population, hosts are no longer under selection to evade compatibility, and the system achieves a stable state with intermediate frequency of sanctions and hosts and symbionts fixed for compatible recognition/singaling alleles.

Finally, if the symbionts are initially fixed for one recognition type (Figure 3, bottom, orange dashed line), hosts rapidly fix for the other recognition type (orange solid line). Once the recognition loci achieve this equilibrium of mutual incompatibility — at which symbiosis is never initiated — host sanctions or symbiont cooperation are no longer under selection. Note, however, that introducing symbionts of the other recognition type (i.e., by mutation) disrupts this equilibrium, and sanctions-cooperation coevolution can then result in fixation of cooperative symbionts (Figure 4).

These individual cases illustrate the balance between coevolution at the recognition loci and the coevolution of host sanctions with symbiont cooperation. If sanctions are less effective (lower ω), there can be complex

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

interactions between coevolution mediated by sanctions and coevolution at recognition/signaling loci. But if sanctions are sufficiently strong (higher ω), they can override this dynamic and maintain symbiont fixation for cooperation. Coevolutionary recursions with different possible values of ω and initial frequency of the symbiont recognition allele p_S illustrate the boundaries of these two conditions (Figure 4). Symbiont cooperation either becomes fixed or lost after 500 generations of coevolution; but hosts may often maintain intermediate frequencies of the sanctioning allele, if the system achieves equilibrium conditions at other loci before sanctions are fixed or lost (Figure 4).

Individual-based simulations

Our analytic models provide good insight into coevolutionary dynamics. However, the approximations necessary to derive our results may limit the models' generality. Also, models of dynamics in a single population do not account for the effects of geographic spatial structure, which can contribute to variation in coevolving species (Nuismer et al. 1999; Thompson 2005; Yoder and Nuismer 2010; Thompson 2013). To examine a broader range of parameter space and the effects of spatial structure, we constructed an individual-based simulation of coevolution between hosts and symbionts in a metapopulation of sites linked by migration. These simulations let us examine the effects of stronger selection, migration and drift (effective population size), and linkage between host and symbiont loci. We ran 1,000 replicate simulations of each of the three host-symbiont models, and 1,000 replicate simulations of neutral evolution, as a standard for comparison (see Methods, below).

Frequency of sanctioning hosts and cooperative symbionts

After 1,000 generations of host-symbiont coevolution, individual-based simulations with host sanctions ended with the host sanctioning allele at significantly higher frequency, averaged across all sites in the metapopulation, than expected from the neutral simulations (Figure 5A; $p < 1 \times 10^{-6}$, one-sided t-test on arcsine-transformed values). At the same time, simulations of sanctions with recognition achieved lower global mean frequency of sanctioning hosts

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

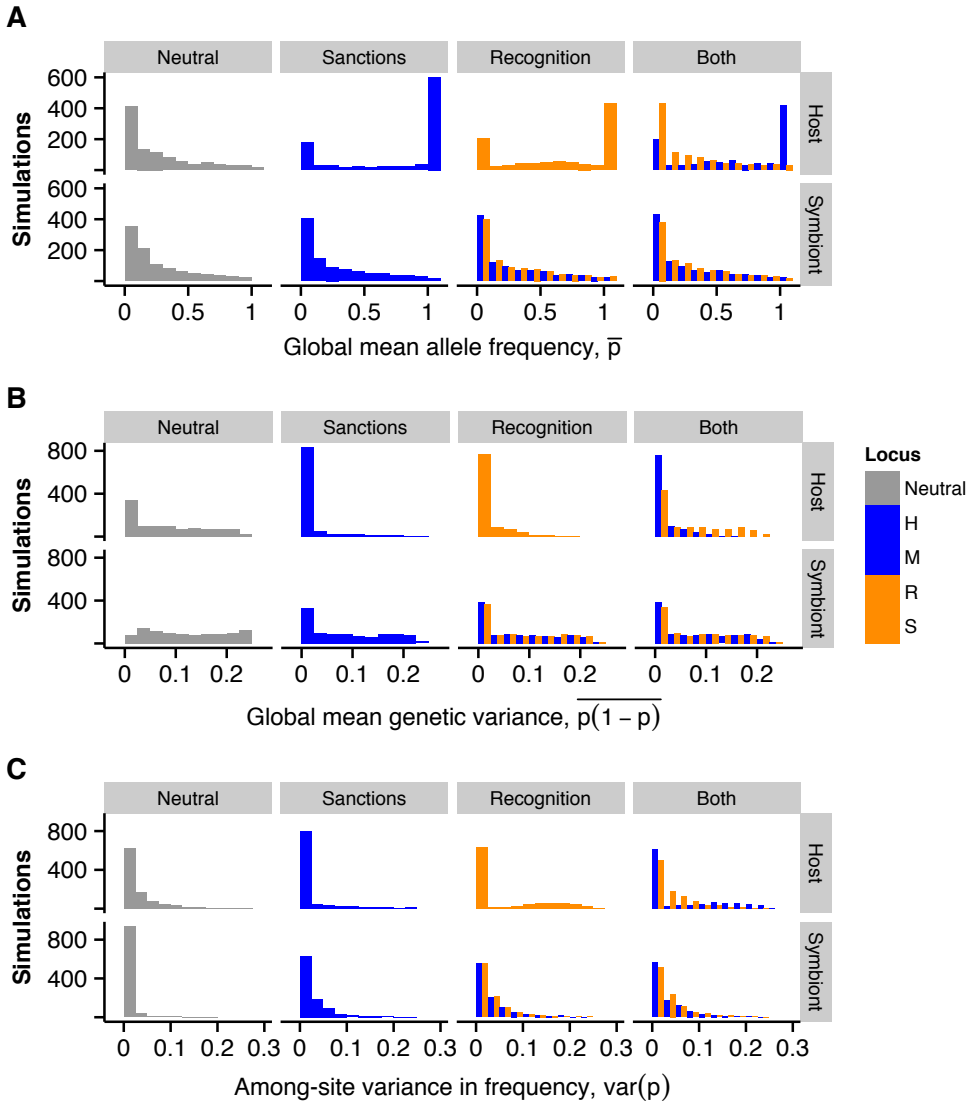

Figure 5: Outcomes of individual-based simulations for each of the three genetic models of symbiosis in a metapopulation of sites linked by migration. Distributions of (A) global mean of within-population allele frequencies, (B) global genetic variation (i.e., expected heterozygosity given global mean allele frequency), and (C) among-site variation in local allele frequency after 1,000 generations of coevolution in 500 replicate simulations run with parameters given in Table 4.

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

than simulations of sanctions alone ($p < 1 \times 10^{-6}$). In simulations of sanctions alone, the global mean frequency of sanctions was strongly negatively correlated with the cost of symbiosis ($\rho = -0.53$, $p < 1 \times 10^{-6}$), and with the benefits of symbiosis to hosts ($\rho = -0.37$, $p < 1 \times 10^{-6}$). By contrast, in simulations of sanctions with recognition the relationships between the cost and benefit of symbiosis and the frequency of sanctions were weak ($\rho = 0.04$, $p = 0.2155$; and $\rho = -0.01$, $p = 0.7752$).

Each of the three models, sanctions and recognition alone maintained cooperative symbionts at higher frequency than expected from the neutral simulations (Figure 5A; t-test $p = 0.01$, 0.05, and 0.07, respectively), though the differences in expected frequency of cooperative symbionts were only marginally greater than seen in the neutral simulations. Cooperative symbionts were generally maintained in the simulated meta-populations, though at low frequency; the median global average frequency of cooperative symbionts was 0.15 in simulations of sanctions alone, 0.16 in simulations of recognition alone, and 0.15 in simulations with both systems. However, in all three models the cooperation alleles were more often lost than alleles in neutral simulations (19% of sanctions-only simulations, 25% of recognition-only, and 25% of simulations with both; compared to 9% of neutral simulations). There was no significant correlation between the frequency of cooperative symbionts and the costs or benefits of symbiosis to symbionts, in any of the three models.

Genetic variation

Calculating global average genetic variance from local allele frequncies, as $p(1-p)$ allows for comparing the fate of genetic variation at host and symbiont loci under the different models of interaction (Figure 5B). For hosts, all coevolutionary scenarios result in significantly lower genetic variance at both sanctions and recognition loci than the neutral expectation ($p < 1 \times 10{-}6$ in one-sided t-tests on arcsine-transformed values from generation 1,000). Variation at host sanctions loci did not differ in simulations of sanctions with or without recognition ($p = 0.98$), but recognition loci maintained more genetic variance in the simulations of sanctions with recognition than with recognition alone ($p < 1 \times 10^{-6}$). In simulations with both sanctions and recognition, the recognition locus maintained significantly greater variance than the sanctions locus ($p < 1 \times 10^{-6}$).

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In symbionts, all three models of symbiosis led to significantly lower genetic variance than seen in neutral simulations (Figure 5B; $p < 1 \times 10^{-6}$ for all comparisons). Variance at the cooperation locus was significantly lower in simulations of sanctions with recognition than in simulations of sanctions alone ($p = 1 \times 10^{-3}$), and in simulations of sanctions with recognition the symbiont cooperation locus maintained lower variance than the recognition locus ($p = 0.05$). However, the results of simulations of sanctions with recognition did not significantly differ from simulations of recognition alone at either the cooperation or the recognition locus.

Geographic variation

The variance in local allele frequencies among sites in the simulated metapopulation also differed among the models (Figure 5C). In hosts, all three models resulted in significantly lower among-site variation than the neutral simulation (in all cases, $p < 1 \times 10^{-6}$ for Wilcoxon sign-rank tests), except for the recognition locus in simulations of sanctions with recognition, which had significantly greater among-site variation ($p = 0.02$). Host sanctions and recognition loci both maintained more among-site variation in allele frequencies under the model of sanctions with recognition than in either system alone ($p < 1 \times 10^{-6}$ in both cases). Finally, in simulations of sanctions with recognition, among-site variation was significantly higher at recognition loci than at sanctions loci ($p < 1 \times 10^{-4}$).

In symbionts, the three models all resulted in greater among-site variation than seen in equivalent neutral simulations (Figure 5C; $p < 1 \times 10^{-6}$ in all cases). Simulations of sanctions resulted in lower among-site variation at symbiont cooperation loci than simulations of recognition ($p = 0.03$) or sanctions with recognition ($p = 0.04$). However, among-site variation at symbiont cooperation loci was not significantly different in simulations of sanctions with recognition than simulations of recognition alone. There was also no difference in among-site variation at signalling loci in simulations of recognition alone versus sanctions with recognition.

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

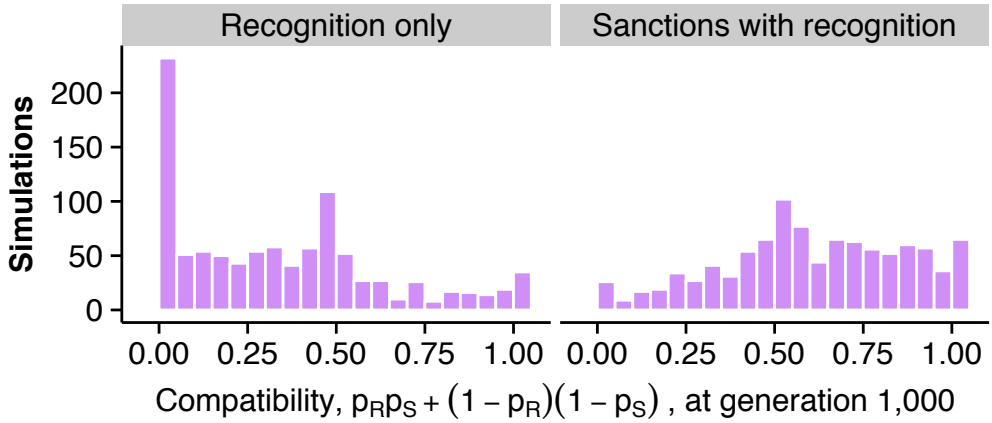

Figure 6: Host-symbiont compatibilty, after 1,000 generations of coevolution in simulations of recognition alone (left) or sanctions with recognition (right).

Coevolution of hosts and symbionts

The global mean frequency of the host sanctions allele was negatively correlated with the frequency of the symbiont cooperation allele in simulations with sanctions alone ($\rho = -0.15$, $p < 1 \times 10^3$) and in simulations of sanctions with recognition ($\rho = -0.48$, $p < 1 \times 10^6$). In contrast, the global mean frequency of the host recognition allele R was negatively correlated with the frequency of the matching symbiont allele in simulations of recognition alone ($\rho = -0.41$, $p < 1 \times 10^{-6}$); but there was no significant association in simulations of sanctions with recognition ($\rho = 0.04$, $p = 0.22$).

Among-site variation in the two species was also often correlated. Greater among-site variation at the host sanctions locus was not significantly associated with greater variation at the symbiont cooperation locus in simulations of sanctions alone ($\rho = 0.03$, $p = 0.28$) — this likely reflects the strong decrease in among-site variation seen at the host sanctions locus in these simulations (Figure 5C). In contrast, among-site variation in host sanctions and symbiont cooperation were strongly correlated in simulations of sanctions with recognition ($\rho = 0.56$, $p < 1 \times 10^{-6}$). Among-site variation at the host recognition locus was strongly correlated with variation at both the symbiont cooperation and recognition loci in simulations of recognition alone ($\rho = 0.57$,

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$p < 1 \times 10^{-6}$ in both cases); but these correlations were much weaker in simulations of sanctions with recognition ($\rho = 0.04$, $p = 0.19$; and $\rho = 0.04$, $p = 0.26$).

Interactions between sanctions/cooperation and recognition

Coevolutionary selection in the simulations of recognition and sanctions with recognition created greater LD between loci, relative to the neutral simulations — but the greater LD was never substantial. For symbionts in the simulations of recognition alone, median LD (estimated as $|D'|$ and averaged across the metapopulation), between the recognition and cooperation loci was 0.002 after 1,000 generations of coevolution, and the greatest observed $|D'|$ was 0.04. In the simulations of sanctions with recognition, median $|D'|$ between host and symbiont loci was similarly low (respectively: median = 0.002, maximum = 0.133; median = 0.002, maximum = 0.047).

Although substantial LD did not emerge in the simulations, the two genetic systems (i.e., sanctions/cooperation and recognition/signaling) did evolve in response to each other, as expected from our analytic models (Figures 3 and 4). From the allele frequencies at the host recognition locus (p_R) and the symbiont signaling locus (p_S), it is possible to calculate the average host-symbiont *compatibility* as $p_R p_S + (1 - p_R)(1 - p_S)$, the probability that a randomly drawn host will initiate symbiosis with a randomly drawn symbiont. In simulations of sanctions with recognitino, host-symbiont compatibility was much higher, after 1,000 generations of coevolution, than in the model of recognition alone (Figure 6, $p < 1 \times 10^{-6}$ in a t-test on arcsine-transformed values).

In simulations of sanctions with recognition, host-symbiont compatibility is strongly and positively correlated with the frequency of sanctioning hosts, p_H (Figure 7, correlation $\rho = 0.19$, $p < 1 \times 10{-}6$).

Discussion

The establishment of successful mutualistic interactions requires communication between potential hosts and symbionts as well as cooperation between

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

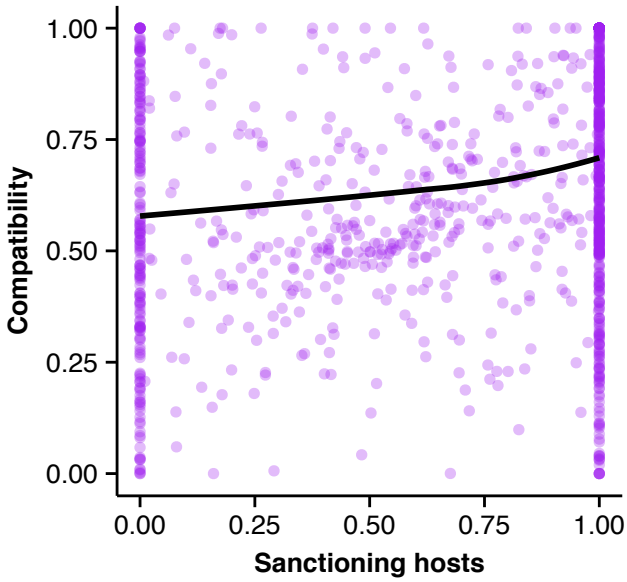

Figure 7: Scatterplot of host-symbiont compatibilty versus the frequency of sanctioning hosts, p_H, after 1,000 generations of coevolution in simulations of sanctions with recognition, with predictions for a linear regression model estimated from arcsine-transformed values. Correlation of arcsine-transformed values is strongly significant ($\rho = 0.20$, $p < 1 \times 10{-}6$).

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

partners once symbiosis is established. Previous models of mutualism evolution have not, however, explicitly included both of these stages. The models we present here reveal that including both stages — communication and cooperation — can result in the maintenance of variation in mutualist quality and promote greater geographic variation in symbiont quality than models that include only communication or cooperation.

The three analytic models we examine are simple, describing only evolution in response to host-symbiont selection, and two of them require substantial simplifying assumptions to find approximate analytical solutions. Nevertheless, the analytical solutions show how host-symbiont selection operates in symbiosis mediated by sanctions, recognition, or both systems together. Even when the fitness effects of mutualism are small and loci responsible for sanctions and recognition are in linkage equilibrium, the evolution of one genetic system influences the evolution of the other, and this can result in maintenance of variation in mutualism quality.

The individual-based simulations explore conditions beyond those assumed in the analytic models, incorporating stronger selection, variable recombination between loci, and the effects of genetic drift and migration across a metapopulation. Their results are nevertheless qualitatively similar to the analytic models: sanctions and recognition systems interact by altering the coevolutionary conditions each genetic system faces (Figure 7). Notably, even with low rates of recombination, coevolution in the simulations does not create substantial LD between the hosts' sanctions and recognition loci, or between symbionts' cooperation and signaling loci. Additionally, the three models produce qualitatively and quantitatively similar expectations for the evolution of host traits but very similar expectations for the evolution of symbiont traits. In fact, for symbionts the allele frequencies predicted by the three models are very similar to those generated by neutral simulations (Figure 5A, 5B). This is consistent with the possibility that, in these simulations, variation in symbiont cooperation is maintained more by mutation-selection-drift balance than by coevolutionary selection.

It is unlikely that different simulation parameters would change this pattern in the symbiont outcomes. Symbiont population sizes and the payout symbionts receive from symbiosis in the simulations are, if anything, smaller than would be realistic for many microbial symbioses, and increasing both in the simulations would make coevolutionary selection stronger and more effective.

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Yet hosts, which have smaller populations and experience weaker fitness effects of symbiosis, show substantial and significant effects of coevolutionary selection on the frequency of sanctioning alleles (Figure 5A, 5B). It seems, then, that the different genetic systems we consider shape the evolution of symbionts' among-site variation, but not the global frequency of cooperation or recognition alleles (Figure 5C).

Comparing the models

In the first, simplest model, host sanctions reliably maintain cooperative symbionts at high frequency. The analytic model of host-symbiont coevolution, which includes only selective dynamics, identifies conditions under which the frequency of cooperative symbionts may decrease, but the loss of cooperation increases the relative fitness of sanctioning hosts — and once sanctions are sufficiently common, the frequency of cooperative symbionts increases to fixation (Figure 1). This recapitulates results widely seen in theory of cooperation between species (Trivers 1971; Axelrod and Hamilton 1981; Bull and Rice 1991; West et al. 2002a; West et al. 2002b; Foster and Wenseleers 2006).

In individual-based simulations of the sanctions-only model, sanctions achieve fixation in the majority of replicate simulations (Figure 5A), resulting in reduced host genetic diversity (Figure 5B) and lower among-site variation in allele frequency (Figure 5C) than seen in neutral simulations. In contrast to the analytical model, simulations of sanctions alone do not result in higher frequency of cooperative symbionts than expected from the neutral simulations (Figure 5A). However, simulations of sanctions did result in greater among-site variation in the frequency of cooperative symbionts (Figure 5C), suggesting that coevolutionary selection in this system promotes geographic diversity in mutualist quality, if not diversity within individual populations.

Symbiosis mediated by host-symbiont recognition, meanwhile, is unable to maintain cooperative symbionts at non-zero frequency, unless cooperative symbionts are already fixed in the population (Figure 2, lower panel). When cooperative symbionts are lost, selection favors hosts compatible with whatever symbiont recognition allele is least common. That is, hosts that avoid symbiosis have a selective advantage. Since even non-cooperative symbionts obtain the full benefit of symbiosis in this model, loss of cooperative sym-

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bionts creates inverse-frequency-dependent coevolutionary selection, leading to cyclical evolution of allele frequencies at the host recognition locus and the symbiont signaling locus (Figure 2, upper panel). These dynamics would maintain variation in hosts' initiation of symbiosis — much as variation is maintained at host immune recognition loci in antagonistic interactions (Dieckmann et al. 1995; Agrawal and Lively 2002). In fact, the loss of cooperative symbionts would effectively convert the initially mutualistic symbiosis to something better described as host-parasite coevolution.

Results from individual-based simulations largely align with these conclusions. Simulations of symbiosis mediated by recognition alone maintained lower global genetic variation in hosts than seen in neutral simulations, but also more than that maintained at the sanction locus in simulations of sanctions alone or recognition alone (Figure 5B). Simulations of recognition alone maintained less among-site variation than seen in neutral simulations, suggesting that these cycles were synchronized across the metapopulation. For symbionts, this system did not increase the proportion of simulations with fixation of alternate recognition alleles, but it did result in elevated among-site variation, at both recognition and cooperation loci.

In contrast to these simpler systems, when hosts are both able to sanction non-cooperative symbionts and to selectively initiate symbiosis based on compatibility at recognition loci, hosts can maintain variation in their compatibility with symbionts, or their ability to sanction (Figure 3). Depending on the effectiveness of sanctions, initial allele frequencies at host and symbiont recognition loci can determine the outcomes at the host sanctions locus and the symbiont cooperation locus (Figure 4). However, these local dynamics predict that symbiont cooperation should either fix or be lost for most starting conditions.

Introducing the effects of drift and migration across a metapopulation, in the individual-based simulations, maintains variation at symbiont cooperation and recognition loci, though coevolutionary selection tends to result in less variation than neutral evolution (Figure 5). Still, simulations of sanctions with recognition maintained more variation at both host sanctioning and recognition loci than simulations of either system alone; as well as greater among-site variation in both hosts and symbionts (Figure 5C).

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Interactions of coevolving genetic systems

Importantly, the interactions between the sanctions/cooperation loci and the recognition/signaling loci shown here emerge in the absence of substantial LD between host loci or between symbiont loci. Instead, what we see is that coevolution by one genetic system changes the selective environment for the other system. Fixation of incompatible host-symbiont recognition loci, for instance, can allow the sanctions/cooperation system to achieve selective equilibrium at intermediate frequencies — since when no hosts initiate symbiosis there is no advantage to sanctioning, and no cost or benefit to symbionts' cooperation (Figure 3, bottom panels). Contrariwise, fixation of symbiont cooperation or host sanctions can allow selective equilibrium at the recognition loci (Figure 3, top panel). Finally, if symbionts fix for cooperation and hosts and symbionts achieve fixation for compatible recongition alleles, hosts may settle at intermediate frequency for sanctions (Figure 3, middle panel).

These equilibria are unstable in the approximated analytic model we analyze, but the results of individual-based simulations show that similar results can emerge and persist in the face of mutation, drift, and migration. In particular, coevolution at the sanctions locus can help to maintain a higher degree of host-symbiont compatibility, as sanctioning hosts experience less risk from initiating symbiosis. This is seen both in comparing host-symbiont compatibility in simulations of recognition alone to simulations of sanctions with recognition (Figure 6), and among replicate simulations with both systems (Figure 7).

Conclusions

An important class of mutualism models, based on economic contract theory, have suggested that sanctioning mechanisms are best understood not as special adapations for minimizing the cost of hosting non-cooperative symbionts (Weyl et al. 2010; Archetti et al. 2011a; Archetti et al. 2011b). Instead, they propose that sanctions are often pre-existing features of the hosts which provide "partner fidelity feedback" by positively responding to cooperative symbionts, or failing to produce rewards for non-cooperative ones (Archetti et al. 2011a). Coevolution in the multi-locus model of symbiosis we present here converges on a result similar to those from economic contact theory.

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Interaction with symbionts of varying quality favors higher frequency, and often fixation, of sanctioning hosts (Figure 3, 4, 5). In turn, sufficiently high frequency of sanctions relaxes selection for host recognition alleles that prevent symbiosis, leading to higher host-symbiont compatibility (Figure 6, 7). In other words, when hosts sanction at sufficient frequency, they are less likely to exclude symbionts using the recognition system. This recapitulates the result of Archetti *et al.*, (2011a) that hosts offering the right "terms" to symbionts need not screen for cooperative symbionts prior to initiating interaction. (In our model, the "terms" offered to symbionts are for cooperation in ongoing interactions.)

An observation supporting partner fidelity feedback models is that sanctioning mechanisms may often be derived from, or identical to, host adaptations to stress or habitat quality independent of mutualism. In obligate pollination mutualism, floral abortion in response to pollinator overexploitation may be a repurposing of plant responses to floral damage; legumes' reduced allocation to unproductive root nodules may arise from adaptations for root growth in heterogenous soil (Weyl et al. 2010). This biological scenario aligns well with the results for sanctions with recognition. If the means to apply sanctions pre-date a mutualism, sanctioning responses would likely be at high frequency from the origins of the interaction, and coevolution with symbionts will maintain a high degree of "willingness" to initiate in symbiosis — for instance, compatible variants at receptors that might otherwise trigger an immune response against symbionts.

This conceptual link between partner fidelity feedback models and earlier models of mutualism is one of several advantages of modelling mutualism as mediated by two separate genetic systems. As in models of host-parasite interaction (Dybdahl et al. 2014), considering multiple forms of host-symbiont interaction interacting simultaneously reveals a picture of mutualism that better captures the complexity of empirical systems (Triplett and Sadowsky 1992; Svensson et al. 2005; Edwards et al. 2006; Heath 2009; Soler et al. 2011), and offers a partial explanation for the paradox that has emerged from theory and empirical study — that contemporary populations of mutualists often exhbibit considerable genetic variation for mutualism performance, yet these interactions are stable over evolutionary time (Heath and Stinchcombe 2013).

Future empirical study of mutualism must anticipate that multiple genetic

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

systems, which may experience very different forms of coevolutionary selection, can contribute to the outcomes of cooperative interactions between species. Research in the legume-rhizobium symbiosis, the mutualism in which specific causal genes are best understood, already provides an example for studies of other systems, through experimental manipulations that account for recognition and sanctioning responses separately (Stanton-Geddes et al. 2013; e.g., Regus et al. 2014) and using population genomics data to differentiate candidate genes experiencing different forms of selection (e.g., Paape et al. 2013; Bonhomme et al. 2015). For many well-studied mutualisms, the genetic resources necessary for the latter are still in the earliest stages of development. These resources must be built with the expectation that the genetic systems mediating mutualism can be more complex than we have previously thought.

Methods

Analytic models

Full details of the derivation of our analytic models of host and symbiont coevolution, and evaluation of equilibrium conditions in those models, are available as Mathematica notebooks provided at github.com/jbyoder/mutualism-sanctions-recognition.

Individual-based simulations

Parameters for our individual-based simulations are listed in Table 4. The simulation script (available at github.com/jbyoder/mutualism-sanctions-recognition), which runs in R (R Core Team 2015), creates N populations of K_i haploid individuals for each species i. We chose parameters ensuring that the interaction would be commnesal or mutualistic (all $B_i \geq C_i$); and also that symbionts would usually have larger population sizes than hosts and experience greater benefits from symbiosis — asymmetries that are seen in many natural mutualistic symbioses.

The simulation starts by randomly creating individuals' genotypes of one or two loci, depending on the model simulated, based on starting allele frequencies drawn from an approximation of the allele frequency spectrum

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 4: Parameter values for the individual-based simulations.

Parameter[1]	Host	Symbiont
For metapopulation structure		
N, number of sites	50	50
K, per-site population size	$U(20, 200)$	$U(200, 2000)$
m, among-site migration rate	$U(0, 0.05)$	$U(0, 0.05)$
For interaction payouts		
C, cost of symbiosis	$U(0.01, 0.5)$	$U(0.01, 0.5)$
B, benefit of symbiosis	$C_H \times U(1, 10)$	$C_S \times U(1, 100)$
ω, effectiveness of sanctions	$U(0.1, 0.9)$	$U(0.1, 0.9)$
For genetics		
r, recombination rate	$U(0, 0.05)$	$U(0, 0.05)$
μ, mutation rate	10^{-6}	10^{-6}

[1] Parameters are either point values, or drawn from a uniform distribution with range $U(min, max)$.

for a standard neutral coalescent model at equilibrium (Ganapathy and Uyenoyama 2009). After creation of the starting populations, the simulation proceeds through the following generational cycle:

Mating. Mating occurs between pairs of hermaphroditic individuals of each species, drawn at random from the same population, with replacement. Each mating produces one offspring, with genotypes at each locus drawn from the parental genotypes. In two-locus species, recombination between loci occurs with probability r_i, and mutation from one allele to the alternate allele occurs with probability μ_i for each locus. The simulation script draws mated pairs until K_i offspring are created, then replaces the parental population with those offspring.

Migration. A proportion m_i of the individuals in each population are selected at random to join a global migrant pool, which are then distributed at random back among the N populations.

Coevolutionary selection. Within each population, hosts interact with randomly-drawn symbionts, with each individual's fitness outcome from the

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

interaction determined by their genotype and the genotype of their host or symbiote, following one of the three models outlined above. All hosts interact; symbionts that are not drawn for interaction with a host are lost from the population. Interacting individuals of both species survive to be used as parents for the next generation if their fitness outcome from the interaction exceed a normally distributed random value; an additional 5% of individuals are randomly drawn to survive in each population, preventing local extinctions.

We ran 1,000 replicate individual-based simulations for each of the three genetic models of host-symbiont mutualism, as well as 1,000 replicate simulations of host and symbiont evolution with no coevolutionary selection, to provide a neutral expectation of results from drift and migration across the metapopulation. Simulation parameters were either held constant across all replicates or drawn at random from a uniform distribution of reasonable values (Table 4).

Acknowledgements

We thank attendees of the UBC Biodiversity Lunchtime Informal Seminar Series for their thoughts on a preliminary analysis of our results, and Scott L. Nuismer for helpful comments on a first draft of this paper. This work was funded by a National Science Foundation award to PT (grant 1237993, N.D. Young, P.I.), and we developed and performed our simulations on computing systems maintained by the Minnesota Supercomputing Institute and Compute Canada's Westgrid network. JBY is currently a postdoctoral fellow with the AdapTree Project, supported by the Genome Canada Large Scale Applied Research Project program with co-funding from Genome BC and the Forest Genetics Council of BC (co-Project Leaders S.N. Aitken and A. Hamann).

References

Agrawal, A. F., and C. Lively. 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evol. Ecol. Res. 4:79–90.

Akçay, E., and E. L. Simms. 2011. Negotiation, sanctions, and context dependency in the legume-Rhizobium mutualism. Am. Nat. 178:1–14.

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Archetti, M., and I. Scheuring. 2011. Coexistence of cooperation and defection in public goods games. Evolution 65:1140–1148.

Archetti, M., I. Scheuring, M. Hoffman, M. E. Frederickson, N. E. Pierce, and D. W. Yu. 2011a. Economic game theory for mutualism and cooperation. Ecol. Lett. 14:1300–12.

Archetti, M., F. Úbeda, D. Fudenberg, J. Green, N. E. Pierce, D. W. Yu, and F. Ubeda. 2011b. Let the right one in: A microeconomic approach to partner choice in mutualisms. Am. Nat. 177:75–85.

Axelrod, R., and W. Hamilton. 1981. The evolution of cooperation. Science 211:1390–1396.

Barton, N. H., and M. Turelli. 1991. Natural and sexual selection on many loci. Genetics 127:229–255.

Bonhomme, M., S. Boitard, H. San Clemente, B. Dumas, N. D. Young, and C. Jacquet. 2015. Genomic signature of selective sweeps illuminates adaptation of Medicago truncatula to root-associated microorganisms. Mol. Biol. Evol.

Bull, J. J., and W. R. Rice. 1991. Distinguishing mechanisms for the evolution of co-operation. J. Theor. Biol. 149:63–74.

Cho, I., and M. J. Blaser. 2012. The human microbiome: at the interface of health and disease. Nat. Rev. Genet. 13:260–70.

Cullender, T. C., B. Chassaing, A. Janzon, K. Kumar, C. E. Muller, J. J. Werner, L. T. Angenent, et al. 2013. Innate and adaptive immunity interact to quench microbiome flagellar motility in the gut. Cell Host Microbe 14:571–581.

Dieckmann, U., P. Marrow, and R. Law. 1995. Evolutionary cycling in predator-prey interactions: population dynamics and the red queen. J. Theor. Biol. 176:91–102.

Doebeli, M., and N. Knowlton. 1998. The evolution of interspecific mutualisms. Proc. Natl. Acad. Sci. 95:1–11.

Dybdahl, M. F., C. E. Jenkins, and S. L. Nuismer. 2014. Identifying the molecular basis of host-parasite coevolution: merging models and mechanisms. Am. Nat. 184:1–13.

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Edwards, D. P., R. Arauco, M. Hassall, W. J. Sutherland, K. Chamberlain, L. J. Wadhams, and D. W. Yu. 2007. Protection in an ant-plant mutualism: an adaptation or a sensory trap? Anim. Behav. 74:377–385.

Edwards, D. P., M. Hassall, W. J. Sutherland, and D. W. Yu. 2006. Selection for protection in an ant-plant mutualism: host sanctions, host modularity, and the principal-agent game. Proc. R. Soc. B 273:595–602.

Fishman, J. A. 2015. Immune reconstitution syndrome: how do we "tolerate" our microbiome? Clin. Infect. Dis. 60:45–47.

Foster, K. R., and T. Wenseleers. 2006. A general model for the evolution of mutualisms. J. Evol. Biol. 19:1283–93.

Foster, K. R., T. Wenseleers, and F. L. W. Ratnieks. 2006. Kin selection is the key to altruism. Trends Ecol. Evol. 21:57–60.

Ganapathy, G., and M. K. Uyenoyama. 2009. Site frequency spectra from genomic SNP surveys. Theor. Popul. Biol. 75:346–354.

Heath, K. D. 2009. Intergenomic epistasis and coevolutionary constraint in plants and rhizobia. Evolution 64:1446–1458.

Heath, K. D., and J. R. Stinchcombe. 2013. Explaining mutualism variation: A new evolutionary paradox? Evolution 68:309–317.

Heath, K. D., A. Stock, and J. Stinchcombe. 2010. Mutualism variation in the nodulation response to nitrate. J. Evol. Biol. 23:2494–2500.

Herre, E. A., and S. A. West. 1997. Conflict of interest in a mutualism: documenting the elusive fig wasp-seed trade-off. Proc. R. Soc. B 264:1501–1507.

Hoeksema, J. D. 2010. Ongoing coevolution in mycorrhizal interactions. New Phytol. 187:286–300.

Holland, J. N., T. H. Fleming, N. Sep, S. Url, and T. E. Society. 1999. Mutualistic interactions between Upiga virescens (Pyralidae), a pollinating seed-consumer, and Lophocereus schottii (Cactaceae). Ecology 80:2074–2084.

Jandér, K. C., and E. A. Herre. 2010. Host sanctions and pollinator cheating in the fig tree-fig wasp mutualism. Proc. R. Soc. B.

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Jones, E. I., R. Ferrière, and J. L. Bronstein. 2009. Eco-evolutionary dynamics of mutualists and exploiters. Am. Nat. 174:780–794.

Kiers, E. T., R. a Rousseau, S. A. West, and R. F. Denison. 2003. Host sanctions and the legume-rhizobium mutualism. Nature 425:78–81.

Kiers, E. T., R. A. Rousseau, and R. F. Denison. 2006. Measured sanctions: Legume hosts detect quantitative variation in rhizobium cooperation and punish accordingly. Evol. Ecol. Res. 8:1077–1086.

Kiester, A. R., R. Lande, and D. W. Schemske. 1984. Models of coevolution and speciation in plants and their pollinators. Am. Nat. 124:220–243.

Kirkpatrick, M., T. Johnson, and N. Barton. 2002. General models of multilocus evolution. Genetics 161:1727–1750.

Kopp, M., and S. Gavrilets. 2006. Multilocus genetics and the coevolution of quantitative traits. Evolution 60:1321–1336.

McNamara, J. M., and O. Leimar. 2010. Variation and the response to variation as a basis for successful cooperation. Philos. Trans. R. Soc. B 365:2627–2633.

Mutlu, E. A., A. Keshavarzian, J. Losurdo, G. Swanson, B. Siewe, C. Forsyth, A. French, et al. 2014. A compositional look at the human gastrointestinal microbiome and immune activation parameters in HIV infected subjects. PLoS Pathog. 10.

Ness, J. H., W. F. Morris, and J. L. Bronstein. 2006. Integrating quality and quantity of mutualistic service to contrast ant species protecting Ferocactus wislizeni. Ecology 87:912–921.

Nuismer, S. L., and M. F. Dybdahl. 2016. Quantifying the coevolutionary potential of multi-step immune defenses. Evolution 70:282–295.

Nuismer, S. L., and S. P. Otto. 2005. Host-parasite interactions and the evolution of gene expression. PLoS Biol. 3:e203.

Nuismer, S. L., R. Gomulkiewicz, and B. J. Ridenhour. 2010. When is correlation coevolution? Am. Nat. 175:525–537.

Nuismer, S. L., J. N. Thompson, and R. Gomulkiewicz. 1999. Gene flow and geographically structured coevolution. Proc. R. Soc. B 266:605–609.

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ohtsuki, H. 2010. Evolutionary games in Wright's island model: kin selection meets evolutionary game theory. Evolution 64:3344–53.

Okamoto, T., A. Kawakita, and M. Kato. 2007. Interspecific variation of floral scent composition in Glochidion and its association with host-specific pollinating seed parasite (Epicephala). J. Chem. Ecol. 33:1065–1081.

Paape, T., T. Bataillon, P. Zhou, T. J. Y. Kono, R. Briskine, N. D. Young, P. L. Tiffin, et al. 2013. Selection, genome-wide fitness effects and evolutionary rates in the model legume Medicago truncatula. Mol. Ecol. 22:3525–38.

Pellmyr, O., and C. J. Huth. 1994. Evolutionary stability of mutualism between yuccas and yucca moths. Nature 372:257–260.

Pflughoeft, K. J., and J. Versalovic. 2012. Human microbiome in health and disease. Annu. Rev. Pathol. 7:99–122.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Regus, J. U., K. A. Gano, A. C. Hollowell, and J. L. Sachs. 2014. Efficiency of partner choice and sanctions in Lotus is not altered by nitrogen fertilization. Proc. R. Soc. B 281:20132587.

Sachs, J. L., U. G. Mueller, T. P. Wilcox, J. J. Bull, T. H. E. Evolution, and O. F. Cooperation. 2004. The evolution of cooperation. Q. Rev. Biol. 79:135–160.

Sasaki, A. 2000. Host-parasite coevolution in a multilocus gene-for-gene system. Proc. R. Soc. B 267:2183–8.

Schatz, B., C. Djieto-Lordon, L. Dormont, J. M. Bessière, D. McKey, and R. Blatrix. 2009. A simple non-specific chemical signal mediates defence behaviour in a specialised ant-plant mutualism. Curr. Biol. 19:361–362.

Simms, E. L., and D. L. Taylor. 2002. Partner choice in nitrogen-fixation mutualisms of legumes and rhizobia. Integr. Comp. Biol. 42:369–380.

Soler, C., M. Hossaert-McKey, B. Buatois, J.-M. Bessière, B. Schatz, and M. Proffit. 2011. Geographic variation of floral scent in a highly specialized pollination mutualism. Phytochemistry 72:74–81.

Stanton-Geddes, J., T. Paape, B. Epstein, R. Briskine, J. B. Yoder, J. Mudge, A. K. Bharti, et al. 2013. Candidate genes and genetic architecture of

bioRxiv preprint doi: https://doi.org/10.1101/049593; this version posted April 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in Medicago truncatula. PLoS One 8:e65688.

Svensson, G., M. Hickman Jr, S. Bartram, W. Boland, O. Pellmyr, and R. A. Raguso. 2005. Chemistry and geographic variation of floral scent in Yucca filamentosa (Agavaceae). Am. J. Bot. 92:1624–1631.

Thompson, J. N. 2005. The Geographic Mosaic of Coevolution. University of Chicago Press, Chicago.

Thompson, J. N. 2013. Relentless Evolution. University of Chicago Press, Chicago.

Triplett, E. W., and M. J. Sadowsky. 1992. Genetics of competition for nodulation of legumes. Annu. Rev. Microbiol. 46:399–428.

Trivers, R. 1971. The evolution of reciprocal altruism. Q. Rev. Biol. 46:35–57.

West, S. A., E. T. Kiers, I. Pen, and R. F. Denison. 2002a. Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? J. Evol. Biol. 15:830–837.

West, S. A., E. T. Kiers, E. L. Simms, R. F. Denison, and P. R. S. L. B. 2002b. Sanctions and mutualism stability: why do rhizobia fix nitrogen? Proc. R. Soc. B 269:685–94.

Weyl, E., M. Frederickson, D. Yu, and N. Pierce. 2010. Economic contract theory tests models of mutualism. Proc. Natl. Acad. Sci. 107:15712.

Yang, S., F. Tang, M. Gao, H. B. Krishnan, and H. Zhu. 2010. R gene-controlled host specificity in the legume-rhizobia symbiosis. Proc. Natl. Acad. Sci. 107:18735–18740.

Yoder, J. B., and S. L. Nuismer. 2010. When does coevolution promote diversification? Am. Nat. 176:802–817.

Young, N. D., F. Debellé, G. E. D. G. Oldroyd, R. Geurts, S. B. Cannon, M. K. Udvardi, V. A. Benedito, et al. 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature 480:520–4.